

Assessment of Performing Arts Students' Productions: A Critical Review of an Assessment Model

Richard Isaac Amuah¹ & Latipher Amma Osei²

Abstract

The education enterprise is fraught with diverse strategies for assessing learning outcomes. The diversity manifests attempts to confront the challenges that impede the achievement of objectivity, which is required to enhance the efficacy of evaluating students' learning outcomes. This struggle becomes pronounced as educators engage in the assessment of "practical" performance of art forms that demand a lot of effort to uphold objectivity and meliorate problems relating to reliability and validity that characterize subjective observations made by assessors. This study examined assessment strategies adopted by a department of performing arts with the view to increasing the degree of objectivity associated with the assessment of performing arts "practical" examination. For the study, data was collected from three judges during the end of semester examinations for the courses Ghanaian traditional dance, Dance Techniques and Introduction to Dance Cultures of the World in the 2015/2016 academic year. The results indicated that there was significant difference with an $\alpha = 0.05$, $F(2, 85) = 30.3$, $p = .001$. The analysis of the inter-judge correlation yielded a correlation coefficient of 0.6. This coefficient is fairly weak considering the impact of the assessment on students' future academic development. To ameliorate the perennial problem of subjectivity in performance assessment, the portfolio assessment has been suggested as a compliment to the traditional "panel assessment" of practical performance examination taken by performing arts students of dance.

Introduction

Scholars of education have stressed the importance of student assessment to educational delivery ((Shepherd, 2013; Brookhart, 2011, Dunn & Mulvenon, 2009; Green & Mantz, 2002). Though assessment of students' achievement is viewed as a critical component of the educational systems across the globe the United States of America were among the first countries to institutionalize the assessment of students' achievement (Sabol, 2001). This institution was established at the heels of the successful launching of the Sputnik by the Russians in 1957. The feat, *inter alia*, urged the American government to review her educational system and place emphasis on accountability in educational delivery (Gruber, 2008). To support the implementation of the educational reforms the National Assessment of Educational Achievement (NAEA) was established to gather data in support of the reforms (McMillan, Venable, & Varier, 2013). Considerable attention was accorded this agency and since its inception the literature in educational delivery and assessment has been replete with theories and practice of educational assessment in all fields of study including the arts.

Prior to the establishment of the NAEA, assessment of students' achievement in arts education received little attention in the literature (Gruber and Hobbs, 2002). Perhaps the lack of interest in the field may have been engendered by the position assumed by a school of thought that averred that the arts in education should be viewed as developmental area of study and that it did not require structure in its delivery. These scholars proposed the child-centered approach, which held sway for over four decades. Within the framework of child-centered education, the arts were regarded as a creative endeavor and hence any attempt to assess students' artistic product was viewed as an affront to the development of creative abilities.

¹ Department of Music and Dance, University of Cape Coast, Ghana. E-mail: richardamuah@yahoo.com

² Department of Theatre Arts, University of Education, Winneba, Ghana. E-mail: lammaosei@gmail.com

By the late 1970s assessment in arts education had shifted from the observation of students' creative process to the assessment of students' creative product. Scholars like Barkan, (1962) played a critical role in propagating the new approach to the assessment of students' artistic endeavour. In addition to the traditional approach of observing students' artistic products and evaluating it globally (Boyle & Radocy, 1987), other approaches to the assessment of students' artistic abilities has evolved. These types may be subsumed under the "process criteria" and "product criteria" (Guskey, 1994).

In Ghana, assessment of students' artistic behaviour was introduced at the tertiary level with the advent of the "specialist training school" established at the Prince of Wales college (Tooley, Dixon, & Amuah, 2007). The need to certify teachers who had undergone training as music specialists engendered the adoption of an assessment strategy particularly the product criteria model to collect data on students' academic and artistic achievement. Since then, this approach has been the only source of assessing students' artistic products. The Achimota School for the training of music specialists was moved to Winneba as part of the Specialist Training College in the 1960s (Flolu & Amuah, 2003). The programme was expanded to include music and fine art. The observational model of assessment that involved a panel of judges rating students' artistic product was pursued.

By the close of the 1960s the School of Performing Arts had been established at the University of Ghana. The School offered three programmes comprising music, theatre and dance. As part of the University's certification process students' artistic abilities were assessed through their artistic products and the observational model of assessment was propagated. Departments of music, dance and theatre that were subsequently established by universities in Ghana have adopted this approach.

Post 1960 to 2017

In Ghana, summative assessment has played a very important role in determining students' academic standing. Summative assessments are cumulative evaluations used to measure student growth after instruction and are generally given at the end of a course in order to determine whether long term learning goals have been met. Summative assessment of students' artistic products has been pursued through the observational assessment model. As already discussed, this model allows for a panel of judges to rate students' artistic products. Though this procedure has received criticisms from some scholars (Gruber, 2008; Harlen, 2007; Taras, 2005) it is still pursued as the only means of assessing and evaluating arts students' creative products in Ghanaian universities.

In view of the controversy raised by scholars on the observational assessment model, there is the need to critically review and evaluate it with the view to ameliorating its damaging effect on Ghanaian performing arts students' academic progression. Though this review is long overdue it has not received the requisite attention in the literature and this research sought to gather information to support further discussion on this issue.

This research investigated the inter-rater reliability of panel assessment of students' artistic product and also examined the extent to which "innovative" assessment models could be employed to enhance the evaluation of Ghanaian students' artistic product. Two main research questions guided the study; the first was what is the inter-rater reliability of observational assessment model adopted by a department of performing arts in a university in Ghana and the second one, what contemporary assessment models could be adopted to enhance the assessment and evaluation of Ghanaian performing arts students' artistic product?

This case study reviewed summative assessment data gathered from an end-of-semester final performance of dance students from a university in Ghana. Three sets of data were analyzed to produce the inter-class correlation coefficient. The fully crossed design in which raters assigned scores to all students participating in the summative assessment was adopted for the study. The inter-class correlation (ICC) model, using the version 16 of SPSS software, was adopted for the analysis of the data.

Results

Table 1. Inter-Class Correlation Coefficient for Assessment Data of Three Artistic Productions

Students' Artistic Production	Inter-class correlation Coefficient
Artistic Production 1	.1
Artistic Production 2	.7
Artistic Production 3	.9

Table 2. Table 2: Qualitative Ratings of ICC Values by Cicchetti (1994)

ICC Values	Qualitative Ratings
Less than .40	Poor IRR
Bet. .40 and .59	Fair IRR
Bet. .60 and .74	Good IRR
Bet. .75 and 1	Excellent IRR

Table 2: Qualitative Ratings of ICC Values by Cicchetti (1994)

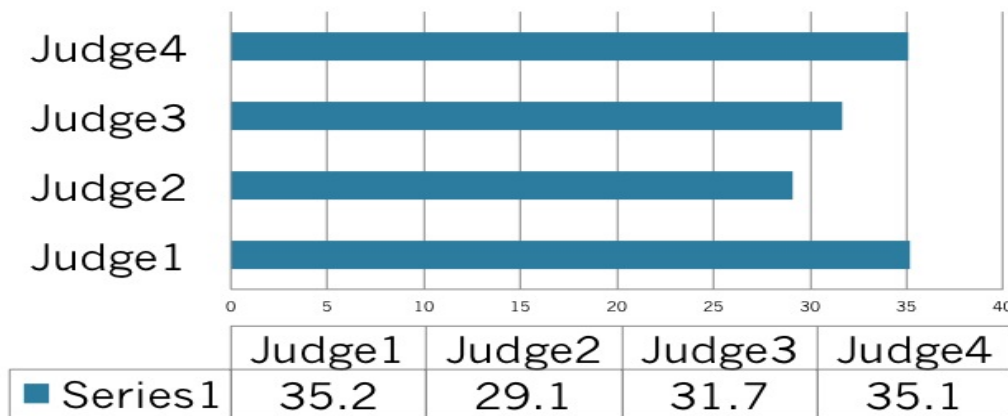
Qualitative ratings of ICC values by Cicchetti (1994) indicate that ICC values less than .40 is a very poor IRR; while ICC values between .40 and .59 are fair IRR; values between .60 and .74 indicate good IRR and lastly ICC values between .75 and 1 are said to be excellent IRR.

Observations

Table 3. Inconsistency of IRR among three sets of data observed

Assessments	Qualitative Ratings
Assessment 1	Poor IRR
Assessment 2	Good IRR
Assessment 3	Excellent IRR

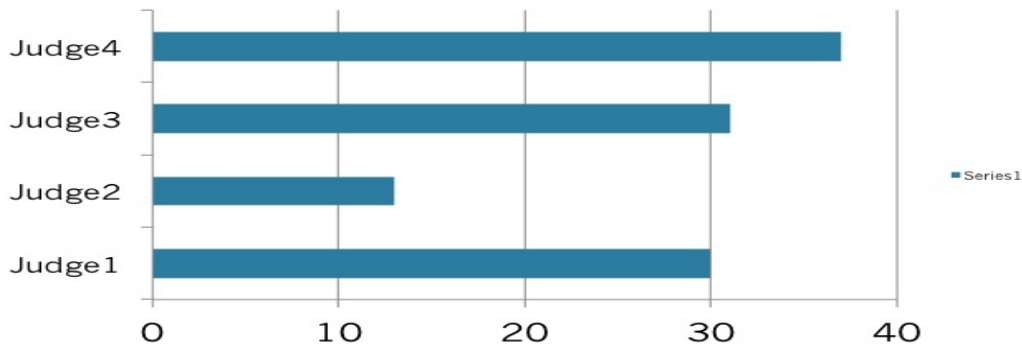
With the above in mind, it was observed that assessment one had an ICC of 0.1 which, indicated a poor IRR; while assessment two which, had an ICC of 0.7 could be rated as having a good IRR and assessment three which had an ICC of 0.9 had an excellent IRR.



Graph 1. Data 1: End-of-Semester Students' Artistic (Dance) Production (Mean Scores)

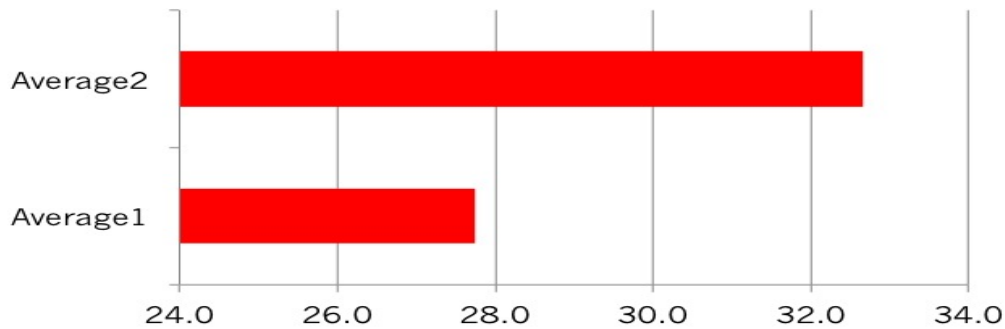
Data One:

As stated earlier, four judges did the grading of the artistic dance productions. With data one, which was one of the three data collected of an end of semester students' artistic dance production, the mean score of judge I was 35.2. Judge II's mean score was 29.1, while judge III's mean score was 31.7 and last but not the least judge IV's mean score was 35.1. With the above, it is obvious to note that judge II's mean score is the outlier.



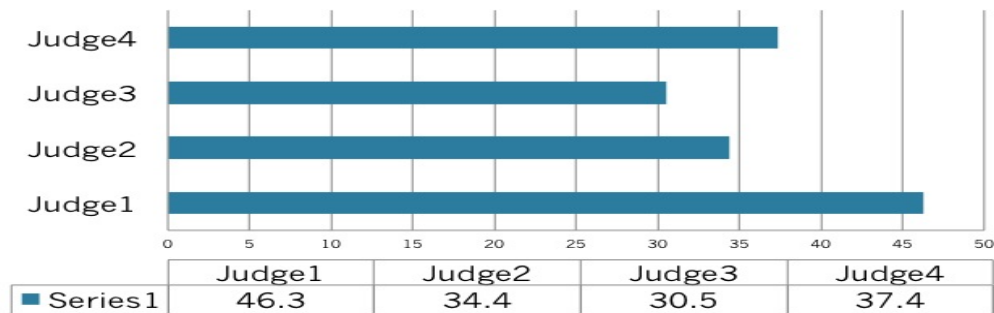
Graph 2: Data 1: End-of-Semester Students’ Artistic (Dance) Production[Individual Student’s Score Per Assessor]

A student was picked to find the scoring per the four assessors/judges. The results indicated that judge I scored the student 30 while judge II scored the same student 13, judge III score the student 31 and judge IV score the same student 37. In this result it is noted that judge II’ score for this student is the outlier.



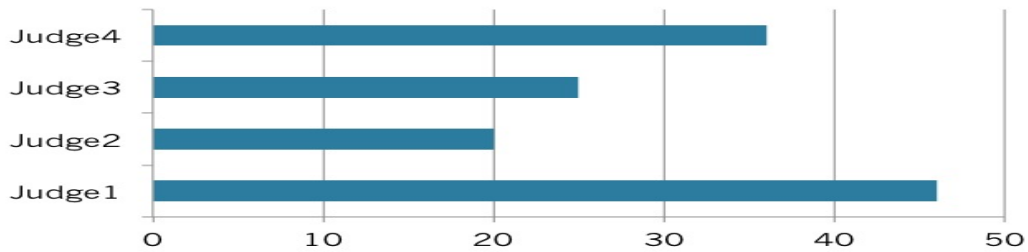
Graph 3. Data 1: End-of-Semester Students’ Artistic (Dance) Production[Mean of Assessors’ Scores Per Student]

When the average mean of this particular student was found of all four judges, it indicated a score of 29. But when the outlier is taken out and the average mean found, this same student scored 32.5.



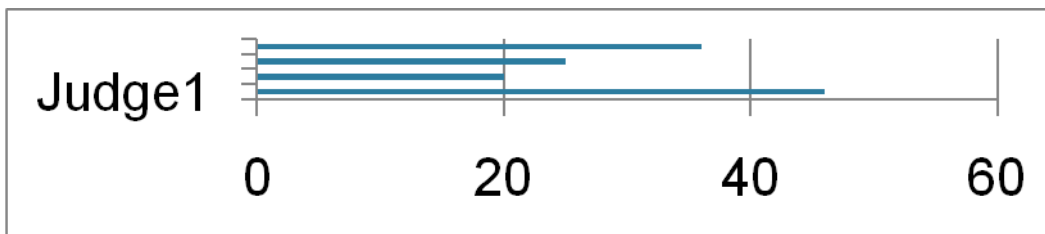
Graph 4. Data 2: End-of-Semester Students’ Artistic (Dance) Production[Mean Scores]

With data 2, which was another of the three data collected, the mean score of judge I was 46.3. Judge II’s mean score was 34.4, while judge III’s mean score was 30.5 and last but not the least judge IV’s means score was 37.4. With the above, it is obvious to note that judge I’s mean score is the outlier.



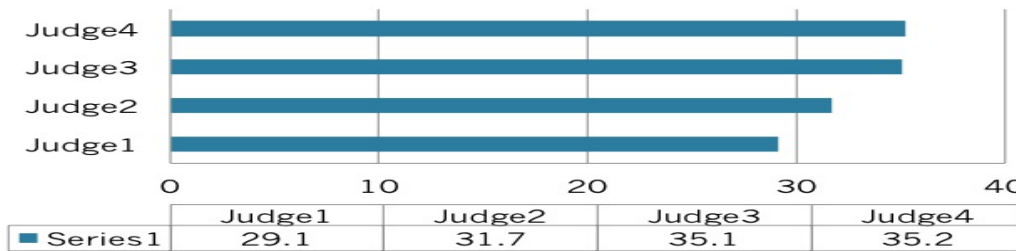
Graph 5. Data 2: End-of-Semester Students’ Artistic (Dance) Production [Individual Student’s Score Per Assessor]

A student was picked to find the scoring per the four judges. The results indicated that judge I scored the student 46 while judge II scored the same student 20, judge III score the student 25 and judge IV score the same student 35. In this result it is noted that judge I’ score for this student is the outlier.



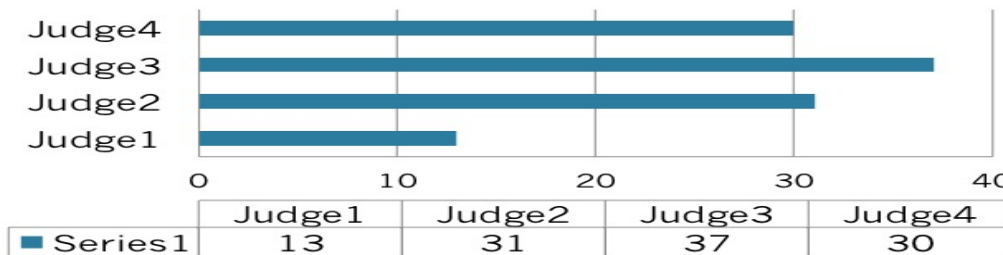
Graph 6. Data 2: End-of-Semester Students’ Artistic (Dance) Production [Mean of Assessors’ Scores Per Student]

When the average mean of this particular student was found of all four judges, it indicated a score of 31.9. But when the outlier is taken out and the average mean found, this same student scored 27.



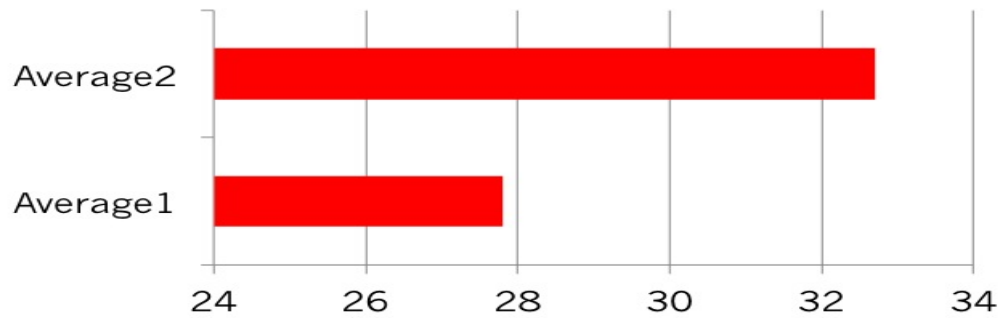
Graph 7. Data 3: End-of-Semester Students’ Artistic (Dance) Production [Mean Scores]

With data two, which was another of the three data collected, the mean score of judge 1 was 29.1. Judge 2’s mean score was 31.7 while judge 3’s mean score was 35.1 and last but not the least judge 4’s means score was 35.2. With the above, it is obvious to note that judge 1’s mean score is the outlier. This is because it is the only one whose mean score is below 30.0 while the other three are all above 30.0.



Graph 8. Data 3: End-of-Semester Students’ Artistic (Dance) Production [Individual Student’s Score Per Assessor]

A student was picked to find the scoring per the four judges. The results indicated that judge I scored the student 13 while judge II scored the same student 31, judge III score the student 37 and judge IV score the same student 30. In this result it is noted that judge I’s score for this student is the outlier.



Graph 9. Data 3: End-of-Semester Students' Artistic (Dance) Production [Mean of Assessors' Scores Per Student]

When the average mean of this particular student was found of all four judges, it indicated a score of 27.9. But when the outlier is taken and the average mean found, this same student scored 32.9.

Discussions

As highlighted by the results of the study the issue of fairness comes into focus when data on students' artistic productions are collected through the application of the observational assessment procedures. It must be noted that this problem is not exclusively peculiar to the assessment of students' artistic products. Educational assessment literature is replete with discussions on fairness in assessment and measurements.

Tierney (2013) avers that, "Fairness is an ideal that has pervaded the history of education in democratic societies, from the establishment of merit-based systems centuries ago to the organization and administration of contemporary formative and summative assessments.

In describing fairness in assessment, Tierney (2013) draws attention to the multiple definitions that could be imposed on the word "fair" and concludes that, "despite what appears to be a multiplicity of definitions, common use of the word fair generally conveys a sense of openness, neutrality, or balance."

By making reference to fairness as an ideal concept, we do not believe that Tierney (2013) is suggesting that we should throw our hands into the air and continue to perpetuate a system that deprives students from obtaining the reward for which they have worked. Rather, there is the need to strive to improve on practices that are detrimental to our students, and the programme we run. It is in this direction that Clarke, Madaus, Horn and Ramos, (2000) suggest a shift of emphasis from summative assessment to formative assessment.

In connection with formative assessment many scholars have suggested the utilization of portfolio assessment as an integral part of formative assessment. This suggestion is founded on what Black and William (2004) refer to as social-constructivist epistemology. Within the purview of this theory, stakeholders particularly students play an active role in the assessment of their artistic products. The portfolio assessment has been used successfully in performing arts education delivery elsewhere. Goolsby (1995) in an article "Portfolio Assessment for Better Evaluation, enumerated a number of schools in the United States that have used the Portfolio Assessment successfully.

In conclusion the data has demonstrated the need for a review of extant assessment practices of the department of performing arts under study as well as other departments pursuing similar assessment procedures. The review could include actions to strengthen existing practices and also exploration and adaptation of contemporary assessment processes including the portfolio assessment procedures.

References

- Barkan, M. (1962). Transition in art education: changing conceptions of curriculum content and teaching. *Art Education*, 15 (7), 12-18.
- Boyle, D. J., & Radocy, R. E. (1987). *Measurement and evaluation of musical experiences*. Schirmer Books.
- Brookhart, S. M. (2011). The use of teacher judgment for summative assessment in the United States: weighed in the balance and (often) found wanting. *Invited Research Seminar Teachers' Judgments within Systems of Summative Assessment: Strategies for Enhancing Consistency*. Oxford University Centre for Educational Assessment.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6 (4), 283-290.
- Clarke, M., Madaus, G. F., Horn, C. L., & Ramos, M. A. (2000). Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum Studies*, 32 (2), 159-181.
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Journal of Practical Assessment, Research and Evaluation*, 14 (7), 1-11.
- Flolu, J., & Amuah, I. (2003). *An introduction to music education in Ghana: For universities and colleges*. Accra: Black Mask Limited.
- Goolsby, T. W. (1995). Portfolio assessment for better evaluation. *Music Educators Journal*, 82 (3), 39-44.
- Green, S. K., & Mantz, M. (2002). Classroom assessment practices: Examining impact on students learning. *Annual meeting of American Educational Research Association*. New Orleans, LA.
- Gruber, D. D. (2008). Measuring student learning in art education. *Art Education*. *Art Education*, 61 (5), 40-45.
- Gruber, D. D., & Hobbs, J. A. (2002). Historical analysis of assessment in art education. *National Art Education Association*, 55 (6), 12-17.
- Guskey, T. R. (1994, 10). Making the grade: what benefits students? *Educational, School, and Counseling Psychology*.
- Harlen, W. (2007). Teachers' summative practices and assessment for learning-tensions and synergies. *The Curriculum Journal*, 16 (2), 207-223.
- McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement: so much more is needed. *Practical Assessment, Research and Evaluation*, 18 (2).
- Sabol, F. R. (2001). Regional findings from a secondary analysis of the 1997 NAEP art assessment based on responses to creating and responding exercises. *Studies in Art Education*, 43 (1), 18-34.
- Shepherd, E. 2. (2013). *Transnational education: The student experience*. Dubai: British Council, Education Intelligence.
- Taras, M. (2005). Assessment - summative and formative - some theoretical reflections. *British Journal of Educational Studies*, 53 (4), 466-478.
- Tierney, R. D. (2013). *Fairness in classroom assessment. Research in classroom assessment*. (M. J. H., Ed.) Los Angeles, CA: SAGE.
- Tooley, J., Dixon, P., & Amuah, I. (2007). Private and public schooling in Ga, Ghana: a census and comparative survey. *International Review of Education*, 53, 3-4.